Computational analysis of thermal stability: effect of Ile→Val mutations in human lysozyme

Yuji Sugita, Akio Kitao and Nobuhiro Go

Background: Free energy calculations are carried out to study the change of thermal stability caused by Ile23→Val, Ile56→Val, Ile89→Val and Ile106→Val mutations in human lysozyme. In order to examine the dependence of the free energy difference, ΔΔG, on the denatured-state structure, extended and native-like conformations are employed as initial conformations in the denatured-state simulations.

Results: Calculated values of ΔΔG for the mutations, Ile56→Val, Ile89→Val and Ile106→Val, were in good agreement with experimental values when the native-like structure was employed in the respective denatured-state simulations. In the case of Ile23→Val, a considerable difference between the calculated and experimental values of ΔΔG was observed.

Conclusions: The physical nature of Ile56→Val, Ile89→Val and Ile106→Val mutations was rationally characterized by a free energy component analysis. It is suggested that the α domain in which Ile23 is included is considerably structured even in the denatured state.

Address: Department of Chemistry, Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan.

Correspondence: Nobuhiro Go
E-mail: go@qchem.kuchem.kyoto-u.ac.jp

Key words: denatured-state structure, free energy calculation, human lysozyme, hydrophobic interaction, thermal stability

Received: **22 December 1997**
Revisions requested: **10 February 1998**
Revisions received: **23 February 1998**
Accepted: **25 February 1998**

Published: **30 March 1998**
http://biomednet.com/elecref/1359027800300173

Folding & Design 30 March 1998, **3**:173–181

Introduction

It is widely believed that the hydrophobic interaction is a major factor stabilizing native protein structures [1]. The contribution of the hydrophobic residues to protein stability has been studied using site-directed mutagenesis followed by thermodynamic and structural measurements [2]. In these studies, several empirical rules have been proposed to explain the free energy differences, ΔΔG, caused by mutations at different locations. Serrano *et al.* [3] have found a statistical correlation between ΔΔG and the 'packing density' (the number of methyl and methylene groups within a 6 Å radius sphere surrounding the deleted methylene group in the native-state barnase). In T4 lysozyme, ΔΔG has been explained in terms of the hydrophobicity of the sidechain and the 'cavity volume' created by mutations [4]. There are also cases, however, in which such a statistical correlation has not been observed [5,6]. This means that the effect of the hydrophobic interaction on protein stability is not explained perfectly by such simple rules.

The change in stability can be enhanced by increasing the stability of the native state, by decreasing the stability of the denatured state, or by a combination of both [2]. Most of the empirical rules are based only on the structural changes in the native state, however. This is understandable because the structures of the denatured state are not well characterized experimentally. Because ΔΔG resulting from a mutation is defined as the difference between the free energy change in the native state, ΔG_N, and the free energy change in the denatured state, ΔG_D, these explanations are valid only when the stability of the denatured state is not changed significantly by the mutation. In the recent experimental studies on the denatured state, it has been shown that the residual structure exists to some extent even in the denatured state and that the structure becomes more compact as the denaturation conditions become milder [7–14]. In these cases, ΔG_D is expected to differ depending on the locations of the mutation sites.

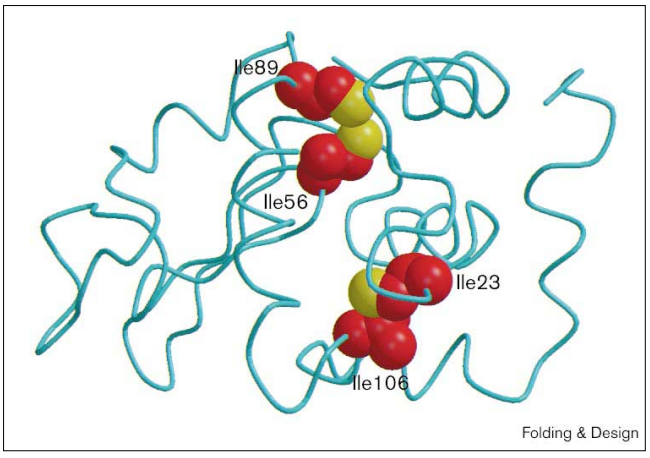
In our previous paper [15], we proposed a computational strategy and actually applied it to investigate protein stability. The effect of the mutation on protein stability should be described based on structures both in the native and in the denatured states. Dissecting the free energy difference into components, we specified the types of interactions (Lennard–Jones, electrostatic, etc.) and the residues that dominantly contribute to ΔΔG.

The method of free energy calculation is based on molecular dynamics (MD; [16–19]). In this method, free energy calculations are performed for both the native and denatured states. In the native-state simulation, a structure determined by X-ray crystallography is employed as an initial structure. On the other hand, an initial structure for the denatured state simulation must be assumed because no such detailed information is available experimentally. A short peptide model, typically a five-residue peptide of the extended structure, including the mutation site in the middle has often been used in the denatured state [20,21].

In our previous study of the Ile56→Val mutation in human lysozyme, five different structures — extended, native-like and three random-coil-like structures — were employed as initial structures of the five-residue peptide in the denatured-state simulations. It was shown that $\Delta\Delta G^{cal}$ significantly depends on the structure of the denatured state and that $\Delta\Delta G^{cal}$ is in good agreement with the experimental free energy difference, $\Delta\Delta G^{exp}$, only when the native-like structure is employed [15]. This suggests that around Ile56 the structure takes a native-like conformation even in the denatured state.

Our purpose in this paper is to calculate free energy differences, $\Delta\Delta G^{cal}$, caused by mutations of a hydrophobic residue at four different sites and to explain the effect each mutation has on the stability in terms of a detailed energetic analysis. We examine the contribution of the denatured state to protein stability using two different structures, the extended and the native-like structures. Human lysozyme contains five isoleucine residues (Ile23, Ile56, Ile59, Ile89 and Ile106), which are located in the hydrophobic core. The thermodynamic and structural changes caused by substitution of these five isoleucine residues by valine (Ile23→Val, Ile56→Val, Ile59→Val, Ile89→Val and Ile106→Val) have been reported already [6]. Here we study four mutations (Ile23→Val, Ile56→Val, Ile89→Val and Ile106→Val; Figure 1). From X-ray crystallographic studies of Ile59→Val in the native state, a new water molecule has been found in the cavity created by the mutation [6]. Thus, the change in stability caused by this mutation originates from insertion of the water molecule as well as the substitution of isoleucine by valine. Because we plan to study the effect of water insertion separately, we focus on the other four mutations in the present study.

Figure 1



The native-state structure of human lysozyme. Heavy atoms of the mutation sites Ile23, Ile56, Ile89 and Ile106 (red) are shown in space filling model format. The mutated methyl groups are shown in yellow. The figure was created using MOLSCRIPT [39] and RASTER3D [40].

Results

The models of the denatured state

Two different models were employed for the denatured-state simulation. These models were chosen carefully according to the strategy that we proposed in a previous paper [15]. In the first model, which is termed the extended model (EXT), we employed a short peptide including the mutation site in the middle. As an initial conformation of this peptide, an extended conformation is used. The second model, which is termed the 'native-like' model (NTV), corresponds to the case in which a local structure around the mutation site is assumed to persist even in the denatured state. The initial coordinates of the peptide were taken from the X-ray structure of the corresponding regions. Even though we did not add any constraints for EXT, the peptide fluctuated around the initial extended structure during the simulation.

When free energy calculations are carried out for a peptide of the extended conformation, it is frequently seen that only the next neighbor and the next two neighbor residues contribute significantly to the free energy change and contributions from outer residues are negligible. This means that a five-residue peptide including the mutation site in the middle is sufficient for EXT. Five-residue peptides are therefore employed in EXT simulations. In NTV, outer residues along the sequence, which are not included in the five-residue peptide, may contribute to the free energy change if the structure around the mutation site takes a specific secondary structure, such as an α helix. Thus, we first calculated the free energy change in the native state, ΔG_N. By considering the contribution from each residue to ΔG_N, the number of residues to be included in NTV was determined. In Table 1, the number of residues, the sequences and the secondary structural code of the corresponding region in the native state employed in NTV are listed. In the cases of Ile23→Val and Ile89→Val, α helices ('B helix' and 'C helix', respectively) start from the residues next to the mutation sites. As a result of considering the results of the free energy calculation in the native state, a nine-residue peptide was employed in the case of Ile23→Val. For the other mutants, five-residue peptides were employed.

Table 1

The native-like models for the denatured-state simulations.

Mutation	Number of residues	Sequence*	Secondary structure†
Ile23→Val	9	ACE-YRGISLANW-NME	BTTB_HHHH
Ile56→Val	5	ACE-YGIFQ-NME	ETTTT
Ile89→Val	5	ACE-DNIAD-NME	SS_HH
Ile106→Val	5	ACE-QGIRA-NME	TGGGG

*ACE, acetyl group; NME, –NH–CH_3 group. †Secondary structural codes of the corresponding region in the native state are determined by the program DSSP [41]. B, bridge; T, turn; H, helix; S, bend; E, sheet; G, 3_{10} helix; and _, other.

Comparison of the calculated free energy changes with the experimental values

Free energy differences between wild-type and mutant proteins, $\Delta\Delta G$, are calculated in free energy perturbation (FEP) and thermodynamic integration (TI) methods [16–19]. In its applications to the problems of protein stability, free energy changes in the native state (ΔG_N) and in the denatured state (ΔG_D) are calculated, then $\Delta\Delta G^{cal}$ is obtained using the relation $\Delta\Delta G^{cal} = \Delta G_N - \Delta G_D$ [21–25].

In Table 2, ΔG_N, ΔG_D and $\Delta\Delta G$ obtained from FEP and TI are listed. Considering the following two criteria, it is judged that ΔG_N and ΔG_D have converged well. First, if differences between ΔG obtained by FEP and TI are negligible, free energy changes are considered to have converged [15]. These differences are < 0.2 kcal/mol in all cases. Second, hysteresis errors (the values after '±') are reasonably small in all cases. For Ile89→Val and Ile106→Val, the hysteresis errors of ΔG_N are relatively large. Thus, the hysteresis errors of $\Delta\Delta G^{EXT}$ and $\Delta\Delta G^{NTV}$ for Ile89→Val and Ile106→Val are also relatively large compared with the other two mutants.

Free energy changes in the native state, ΔG_N, are all negative and are within the range –1.18 kcal/mol to –0.02 kcal/mol. In the denatured state, ΔG_D^{NTV} and ΔG_D^{EXT} are also negative and take the values within the ranges –1.83 kcal/mol to –1.01 kcal/mol and –2.41 kcal/mol to –1.71 kcal/mol, respectively. Because the magnitude of the changes in the denatured state is much greater that that in the native state, $\Delta\Delta G^{NTV}$ ($= \Delta G_N - \Delta G_D^{NTV}$) and $\Delta\Delta G^{EXT}$ ($= \Delta GN - \Delta G_D^{EXT}$) are positive. In other words, mutants are less stable than wild-type lysozyme in all cases.

The correlation between the calculated and experimental free energy differences are shown in Figure 2. $\Delta\Delta G^{NTV}$ for Ile56→Val, Ile89→Val and Ile106→Val are, remarkably, in good agreement with $\Delta\Delta G^{exp}$, whereas $\Delta\Delta G^{EXT}$ for these mutants are larger than $\Delta\Delta G^{exp}$. Thus, NTV is an appropriate model of the denatured state in these three cases. In other words, it is strongly suggested that local structures around the mutation sites, Ile56, Ile89 and Ile106, in the denatured state are native-like rather than extended. Although values of $\Delta\Delta G^{EXT}$ differ by a large amount from $\Delta\Delta G^{exp}$ in Ile89→Val and Ile106→Val, $\Delta\Delta G^{exp}$ is still in the

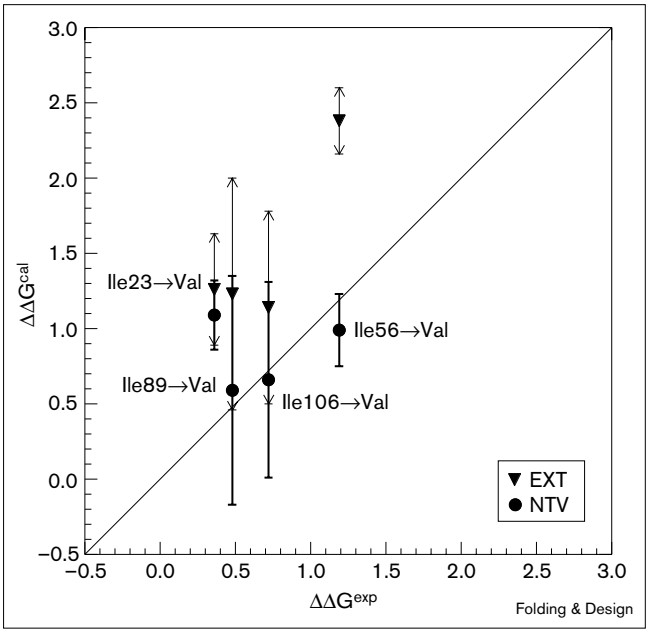
Table 2

Free energy differences and covalent, Lennard–Jones and electrostatic components.

Mutant	Denatured state model	Method	Component	ΔG_N (kcal/mol)*	ΔG_D (kcal/mol)*	$\Delta\Delta G^{cal}$ (kcal/mol)†	$\Delta\Delta G^{exp}$ (kcal/mol)
Ile23→Val	EXT	FEP		−0.75 ± 0.22	−2.00 ± 0.30	1.24 ± 0.38	0.36 ± 0.10
		TI		−0.73 ± 0.23	−2.00 ± 0.29	1.26 ± 0.37	
	NTV	FEP		−0.75 ± 0.22	−1.82 ± 0.01	1.07 ± 0.22	
		TI		−0.73 ± 0.23	−1.83 ± 0.00	1.09 ± 0.23	
			Covalent	0.67 ± 0.09	0.54 ± 0.05	0.14 ± 0.11	
			Lennard–Jones	−1.64 ± 0.14	−2.60 ± 0.04	0.96 ± 0.15	
			Electrostatic	0.23 ± 0.00	0.23 ± 0.01	-0.01 ± 0.01	
Ile56→Val	EXT	FEP		0.00 ± 0.20	−2.39 ± 0.06	2.39 ± 0.21	1.19 ± 0.10
		TI		−0.02 ± 0.21	−2.40 ± 0.06	2.38 ± 0.22	
	NTV	FEP		0.00 ± 0.20	−1.02 ± 0.13	1.02 ± 0.24	
		TI		−0.02 ± 0.21	−1.01 ± 0.12	0.99 ± 0.24	
			Covalent	0.85 ± 0.07	0.47 ± 0.05	0.38 ± 0.09	
			Lennard–Jones	−1.54 ± 0.10	−2.14 ± 0.08	0.60 ± 0.13	
			Electrostatic	0.67 ± 0.04	0.66 ± 0.02	0.01 ± 0.05	
Ile89→Val	EXT	FEP		−1.22 ± 0.79	−2.41 ± 0.12	1.19 ± 0.80	0.48 ± 0.19
		TI		−1.18 ± 0.76	−2.41 ± 0.13	1.23 ± 0.77	
	NTV	FEP		−1.22 ± 0.79	−1.88 ± 0.01	0.66 ± 0.79	
		TI		−1.18 ± 0.76	−1.77 ± 0.01	0.59 ± 0.76	
			Covalent	0.65 ± 0.27	0.38 ± 0.02	0.27 ± 0.27	
			Lennard–Jones	−2.06 ± 0.46	−2.47 ± 0.02	0.41 ± 0.46	
			Electrostatic	0.24 ± 0.03	0.32 ± 0.00	−0.08 ± 0.03	
Ile106→Val	EXT	FEP		−0.57 ± 0.67	−1.64 ± 0.10	1.07 ± 0.67	0.72 ± 0.10
		TI		−0.57 ± 0.64	−1.71 ± 0.09	1.14 ± 0.64	
	NTV	FEP		−0.57 ± 0.67	−1.08 ± 0.05	0.51 ± 0.67	
		TI		−0.57 ± 0.64	−1.23 ± 0.09	0.66 ± 0.65	
			Covalent	0.68 ± 0.27	0.51 ± 0.03	0.17 ± 0.15	
			Lennard–Jones	−1.90 ± 0.55	−2.32 ± 0.12	0.43 ± 0.56	
			Electrostatic	0.65 ± 0.05	0.58 ± 0.00	0.07 ± 0.05	

*Hysteresis errors are $|\Delta G(\text{forward}) - \Delta G(\text{reverse})|/2$. †Hysteresis errors are the standard deviation of four values, $\Delta G_N(\text{forward}) - \Delta G_D(\text{forward})$, $\Delta G_N(\text{forward}) - \Delta G_D(\text{reverse})$, $\Delta G_N(\text{reverse}) - \Delta G_D(\text{forward})$, and $\Delta G_N(\text{reverse}) - \Delta G_D(\text{reverse})$.

Figure 2



The correlation between the calculated and the experimental free energy differences. The calculated free energy differences for the extended (EXT) and native-like (NTV) models are shown with error bars.

range of its error bar. As shown in Table 2, the magnitude of the error is mainly determined by the error in ΔG_N because the errors of $\Delta\Delta G^{EXT}$ and $\Delta\Delta G^{NTV}$ are much smaller than the error of ΔG_N. Thus, in order to assess the difference of EXT and NTV in more detail, the error of ΔG_N should be reduced.

Despite the success of these three cases, the difference between $\Delta\Delta G^{cal}$ and $\Delta\Delta G^{exp}$ is relatively large in Ile23→Val. Because the free energy calculation is well established to give a reasonable free energy value when a structure is determined in atomic detail, we think that ΔG_N is reliable. This indicates that the difference between $\Delta\Delta G^{cal}$ and $\Delta\Delta G^{exp}$ in Ile23→Val is mainly a consequence of the model of the denatured state. The validity of the denatured-state model will be examined in the Discussion section.

Breakdown of the free energy difference into interaction types

In TI, the free energy change ΔG can be decomposed into three components: covalent, ΔG_{cov}; Lennard–Jones, ΔG_{LJ}; and electrostatic, ΔG_{el}. We call the sum of ΔG_{LJ} and ΔG_{el} the nonbonded component (ΔG_{nonbd}).

In Table 2, covalent (ΔG_{cov}, $\Delta\Delta G_{cov}$), Lennard–Jones (ΔG_{LJ}, $\Delta\Delta G_{LJ}$) and electrostatic (ΔG_{el}, $\Delta\Delta G_{el}$) components are listed. There are common features in these free energy components. Covalent and electrostatic components of ΔG_N and ΔG_D^{NTV} are positive, whereas Lennard–Jones components are negative.

All components of $\Delta\Delta G^{NTV}$ in Table 2 are positive or nearly equal to zero, meaning that all components are contributing to destabilize the mutants. The electrostatic contribution $\Delta\Delta G_{el}$ is small in all four cases. The major contribution is from $\Delta\Delta G_{LJ}$ in all four cases, except that $\Delta\Delta G_{cov}$ is also slightly large in Ile56→Val and Ile89→Val. The very small electrostatic contribution is reasonable because the mutation is caused by removing a neutral methyl group. This is by no means trivial in the light of our previous computational study, however [26], in which a valine to an alanine mutation was accompanied by a large $\Delta\Delta G_{el}$ contribution from electrostatic interactions between nearby charged residues. So far we have discussed aspects common to all four mutants.

Nonbonded components (ΔG_{nonbd} and $\Delta\Delta G_{nonbd}$) are divided further into protein–protein ($\Delta G_{pro-pro}$ and $\Delta\Delta G_{pro-pro}$) and protein–water ($\Delta G_{pro-wat}$ and $\Delta\Delta G_{pro-wat}$) components as shown in Table 3. In the case of Ile23→Val, $\Delta\Delta G_{nonbd}$ comes mainly from $\Delta\Delta G_{pro-pro}$, but

Table 3

Nonbonded, protein–protein and protein–water free energy components.

Mutant	Denatured-state model	Component	ΔG_N	ΔG_D	$\Delta\Delta G$
Ile23→Val	NTV	Nonbonded	−1.41 ± 0.14	−2.26 ± 0.33	0.85 ± 0.36
		Protein–protein	−1.58 ± 0.11	−2.20 ± 0.18	0.63 ± 0.21
		Protein–water	0.17 ± 0.03	−0.05 ± 0.15	0.22 ± 0.23
Ile56→Val	NTV	Nonbonded	−0.87 ± 0.14	−1.48 ± 0.06	0.60 ± 0.16
		Protein–protein	−0.94 ± 0.08	−1.27 ± 0.01	0.33 ± 0.08
		Protein–water	0.07 ± 0.07	−0.20 ± 0.08	0.28 ± 0.10
Ile89→Val	NTV	Nonbonded	−1.83 ± 0.49	−2.15 ± 0.01	0.32 ± 0.49
		Protein–protein	−1.91 ± 0.34	−1.84 ± 0.03	−0.08 ± 0.34
		Protein–water	0.09 ± 0.15	−0.31 ± 0.01	0.40 ± 0.15
Ile106→Val	NTV	Nonbonded	−1.25 ± 0.49	−1.74 ± 0.12	0.49 ± 0.51
		Protein–protein	−1.43 ± 0.54	−1.50 ± 0.00	0.07 ± 0.54
		Protein–water	0.18 ± 0.05	−0.24 ± 0.12	0.42 ± 0.13

also non-negligibly from $\Delta\Delta G_{pro-wat}$. In the case of Ile56→Val, $\Delta\Delta G_{pro-pro}$ and $\Delta\Delta G_{pro-wat}$ have contributions of about the same magnitude. In Ile89→Val and Ile106→Val, $\Delta\Delta G_{pro-wat}$ mainly contributes to ΔG_{nonbd}. These different dependencies of $\Delta\Delta G_{pro-wat}$ on the mutation sites should come from different degrees of exposure of the relevant site to solvent water. Table 3 indicates that the exposure to solvent in the denatured state is as important as in the native state.

Residual free energy components

Because $\Delta\Delta G^{NTV}$ for Ile56→Val, Ile89→Val and Ile106→Val are in good agreement with $\Delta\Delta G^{exp}$, we further examine the effect of mutation on protein stability by dividing $\Delta\Delta G^{NTV}$ into the residual free energy components, $\Delta\Delta G_{residue}$. As shown in Table 2, $\Delta\Delta G^{NTV}$ and $\Delta\Delta G^{EXT}$ in Ile23→Val are overestimated. To understand why these values are overestimated, $\Delta\Delta G^{NTV}$ in Ile23→Val is also decomposed into $\Delta\Delta G_{residue}$. In Figure 3, $\Delta\Delta G^{NTV}_{residue}$ values are shown. The residues whose $|\Delta\Delta G_{residue}|$ values are > 0.03 kcal/mol are shown in a ball-and-stick representation in the native structure in Figure 4.

Ile23→Val

For Ile23→Val, considering the residual free energy component of ΔG_N, we employed the nine-residue peptide for the denatured state. As shown in Figure 4a, the residues that contribute significantly to $\Delta\Delta G_{cal}$ are Tyr20, Arg21, Gly22, Ile23, Trp28, Val100 and Trp109. Although Trp28 is the fifth residue from the mutation site along the sequence, it has a significantly large contribution to $\Delta\Delta G_{cal}$. The mutation site Ile23 is on a β turn, which connects two α helices: helix A and helix B. Trp28 is located in the middle of the B helix and is very close in space to Ile23. Ile23→Val will be further discussed later.

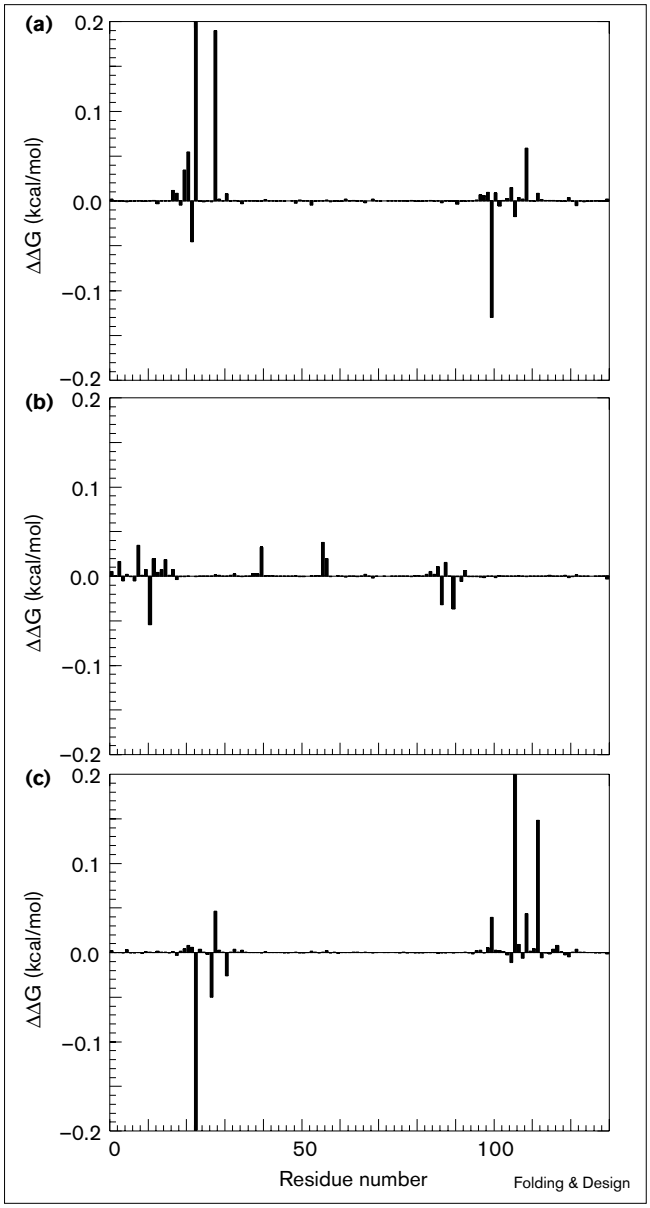
Ile56→Val

We have already reported the detailed analysis of the effect of Ile56→Val on the stability [15]. Thus, here we just point out the residues whose contribution to ΔΔG is significant. All the residues (Phe3, Leu8, Tyr38, Asn39, Thr40 and Ile89) that contribute significantly to $\Delta\Delta G_{cal}$ are located in the hydrophobic core. $\Delta\Delta G_{nonbd}$ from the neighboring region of the mutation site along the sequence was negligible.

Ile89→Val

For Ile89→Val, $\Delta\Delta G_{pro-wat}$ mainly determines ΔG_{nonbd}. Although the total $\Delta\Delta G_{pro-pro}$ is negligible, we show the residues whose $|\Delta\Delta G_{residue}|$ are > 0.03 kcal/mol (Figure 4b). These residues are Leu8, Thr11, Thr40, Ile50, Asp87 and Ala90. They can be classified into two types according to their locations in the native-state structure. The first type involves Leu8, Thr11, Thr40 and Ile56, which are located in the hydrophobic core in which the mutation site, Ile89, is also located. The residues Ile56 and Ile89 face each other in the native state. The residues in the second type are located in the region neighboring the mutation site (Asp87 and Ala90). $\Delta\Delta G_{residue}$ of the mutation site Ile89, is negligible because the Lennard–Jones component of this residue cancels its electrostatic component. $\Delta\Delta G_{nonbd}$ of Leu8, Thr40 and Ile56 are positive, whereas those of Thr11, Asp87 and Ala90 are negative. Thus, the sum of these components are almost zero.

Figure 3

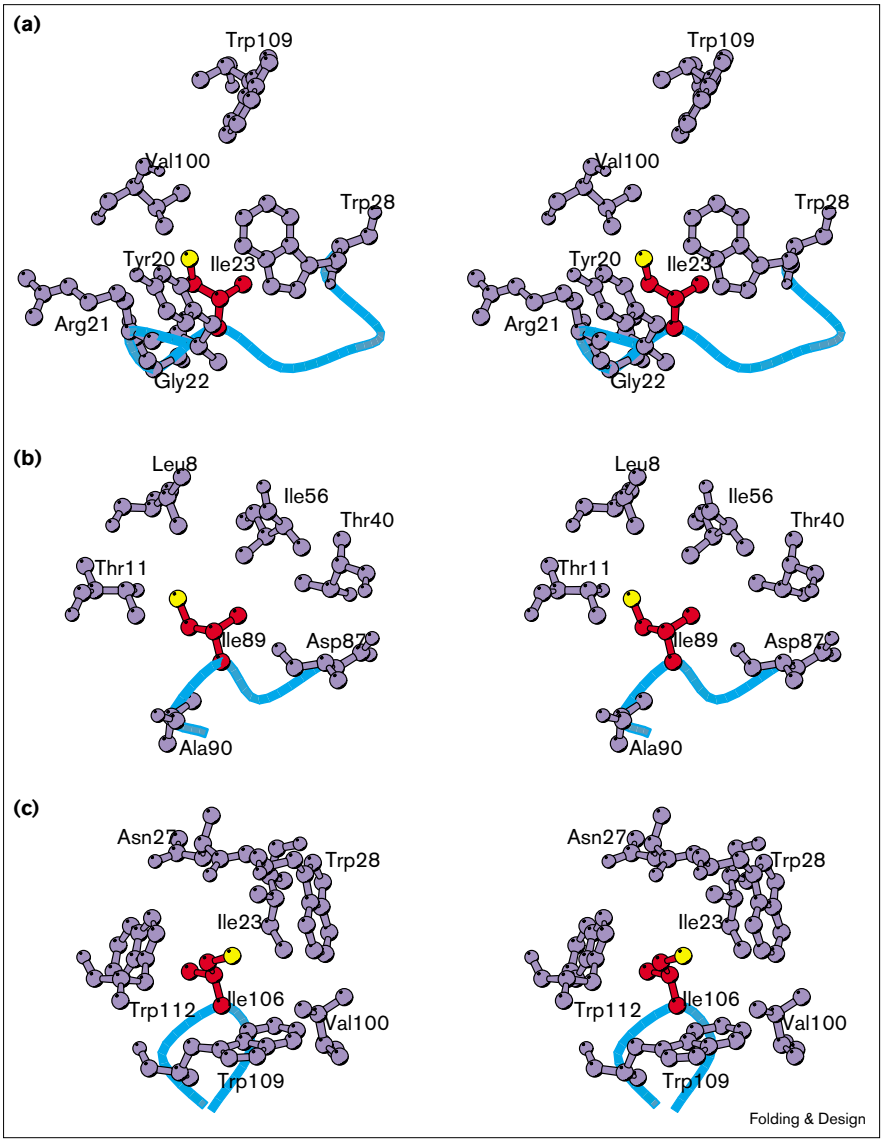


Residual free energy components of $\Delta\Delta G^{NTV}$ for **(a)** Ile23→Val, **(b)** Ile89→Val and **(c)** Ile106→Val.

Ile106→Val

As in the case of Ile89→Val, $\Delta\Delta G_{pro-pro}$ is also negligible in Ile106→Val. The residues, Ile23, Asn27, Trp28, Val100,

Figure 4



Residues whose contributions are significantly large in $\Delta\Delta G_D^{NTV}$ for **(a)** Ile23→Val, **(b)** Ile89→Val and **(c)** Ile106→Val. Heavy atoms of the mutation site (red), the mutated methyl group (yellow), and residues whose $|\Delta\Delta G_{residue}|$ values are > 0.03 kcal/mol (blue). The figure was created using MOLSCRIPT [39].

Trp109 and Trp112, which satisfy the criterion, are shown in Figure 4c. In the native state, Trp112 is located in the middle of the α helix (D helix) close to the mutation site. Thus, Ile106 and Trp112 are very close to each other. Ile23 and Ile106 face each other in the native state. $\Delta\Delta G_{nonbd}$ for the residues from Val100 to Trp112 are positive, whereas $\Delta\Delta G_{nonbd}$ for Ile23 and Asn27 are negative. Thus, the components cancel each other out resulting in an almost negligible value of $\Delta\Delta G_{pro-pro}$.

Discussion

The model of the denatured state in Ile23→Val

Although values of $\Delta\Delta G^{cal}$ determined by NTV are in good agreement with $\Delta\Delta G^{exp}$ in the cases of Ile56→Val, Ile89→Val and Ile106→Val, $\Delta\Delta G^{cal}$ determined either by EXT or by NTV differs significantly from $\Delta\Delta G^{exp}$ for Ile23→Val (Table 2). In this section, we discuss possible reasons for this difference. As we pointed out in the Results section, our free energy calculation is now able to give a reasonable value of free energy when the structure is determined in atomic detail. We therefore think that ΔG_N is reliable. Thus, the disagreement between $\Delta\Delta G^{cal}$ and $\Delta\Delta G^{exp}$ in Ile23→Val should be a consequence of the inappropriateness of the denatured-state model.

Before we discuss the problems encountered when calculating $\Delta\Delta G^{cal}$ for Ile23→Val, we briefly describe the approximation involved with the peptide model, which was detailed in previous papers [15,26]. First, the system is classified into three parts, residues included in the peptide

model (pept), the rest of the residues (rest), and water molecules (wat). Considering the interactions among these three parts, the free energy change can be decomposed into three terms:

$$\Delta G = \Delta G_{pept-pept} + \Delta G_{pept-wat} + \Delta G_{pept-rest} \quad (1)$$

where $\Delta G_{pept-pept}$, $\Delta G_{pept-wat}$ and $\Delta G_{pept-rest}$ are free energy changes caused by the interactions among the peptide part, between the peptide and water molecules, and between residues in the peptide model and the rest of the residues. In the peptide model for the denatured state, $\Delta G_{pept-rest}$ is assumed to be negligible. We therefore point out three possible reasons why we did not succeed in calculating ΔG_D for Ile23→Val. First, the structure of the peptide is not appropriate. In other words, $\Delta G_{pept-pept}$ and $\Delta G_{pept-wat}$ are not calculated correctly. Second, residues that are not included in the peptide model significantly contribute to ΔG (i.e. $\Delta G_{pept-rest}$ is not negligible). Third, a combination of the first and second reasons.

In order to examine these possibilities, we consider the physical picture of the denatured state in lysozyme, which has been obtained from experimental studies. In the recent studies of Dobson and coworkers [27–30], local protein structure in the denatured state has been theoretically modeled using a random coil peptide model, which is based on amino acid specific (ϕ,ψ) distributions extracted from the PDB. The model structure has been compared with the results of NMR measurements. The authors applied their method to hen lysozyme denatured in 8 M urea at low pH. In four regions of the protein, namely residues Asn19–Phe34, Ile58–Cys64, Met105–Arg112 and Asp119–Arg125, marked deviations from random coil predictions were observed for a variety of parameters. This means that these four regions are structured to some extent. The residues employed in the nine-peptide model (Trp20–Trp28), which includes an α helix, overlap the first region pointed out by Dobson and coworkers. Thus, if the first reason above holds true, this region should be structured but not native-like. We think this is less probable than the second reason above by the following reasoning. Lysozyme has two domains — the α domain, which consists of α helices (right-hand side of Figure 1) and the β domain, which mainly consists of β strands (left-hand side of Figure 1). The first, third and forth regions pointed out by Dobson and coworkers (Asn19–Phe34, Met105–Arg112 and Asp119–Arg125) are included in the α domain. Thus, the α domain is expected to remain relatively compact even in the denatured state. It is then possible that some specific peptide–rest interactions make a significant contribution to ΔG_D. In order to understand further Ile23→Val, it is essential to characterize the denatured-state structure.

Analysis of protein stability in terms of the native state and the denatured state structures

$\Delta\Delta G$ depends on the structures of both the native and the denatured states. Most of the empirical rules proposed so far to explain $\Delta\Delta G^{exp}$ are based solely on the native-state structure, however. This situation is understandable because no structural information on the atomic detail is available for the denatured state. Calculation of $\Delta\Delta G$ requires three elements: computational software for the free energy calculation; the structure of the native state; and the structure of the denatured state. The software system we have developed and employed for the free energy calculation has such a high accuracy that calculated $\Delta\Delta G$ can be compared seriously with experimental values as long as structures of the native and the denatured states are reliable. We have a reliable X-ray structure for the native state. The only uncertainty in our scheme of calculation of $\Delta\Delta G$ is in the denatured-state structure. In this situation, we compared the calculated value of $\Delta\Delta G$ ($\Delta\Delta G^{cal}$) with its experimental value ($\Delta\Delta G^{exp}$) in order to deduce the structure of the denatured state.

We have applied this scheme of calculation of $\Delta\Delta G$ for a series of four mutations in human lysozyme. Results are summarized in Table 4. As far as we are aware, this is the first computational attempt to analyze a series of mutations in order to elucidate detailed mechanisms of protein stabilization.

Table 4

Summary of results.

Condition	Ile23→Val	Ile56→Val	Ile89→Val	Ile106→Val
Agreement with experimental value	Fair	Excellent	Excellent	Excellent
Possible structure near mutation site in the denatured state	Not clear in this study	Native-like	Native-like	Native-like
Dominant interactions	Lennard–Jones	Lennard–Jones	Lennard–Jones	Lennard–Jones
Dominant nonbonded components	$\Delta\Delta G_{pro-pro}$	$\Delta\Delta G_{pro-pro}$, $\Delta\Delta G_{pro-wat}$	$\Delta\Delta G_{pro-wat}$	$\Delta\Delta G_{pro-wat}$
Residues significantly contributing to $\Delta\Delta G_{pro-pro}$	**Tyr20**, **Arg21**, **Gly22**, **Ile23**, **Trp28**, Val100, Trp109	Phe3, Leu8, Tyr38, Asn39, Thr40, Ile89	Leu8, Thr11, Thr40, Ile50, **Asp87**, **Ala90**	Ile23, Asn27, Trp28, Val100, Trp109, Trp112

Residues in bold are relatively close to the mutation site along the sequence.

Materials and methods

Simulation procedure

Algorithms of the MD simulations are the same as those we have employed previously [15,26]. Atomic coordinates determined by X-ray crystallography were employed as the initial coordinates of human lysozyme, 196 crystal water molecules, and one sodium ion in the native state simulation. An additional 4484 water molecules were placed around the protein to fill a sphere of 34 Å radius. In the denatured state simulations, either a five-residue or nine-residue peptide was solvated by water molecules to fill a sphere of 20 Å. In both the native and the denatured states, there are approximately three or more layers of water molecules.

The AMBER all-atom energy function [31] was used for the protein and peptide molecules, and the TIP3P model [32] was employed for water molecules. Spherical solvent boundary potential (SSBP; [33]) was applied to the system and cell multipole method (CMM; [34]) was employed to evaluate nonbonded interactions with the sphere. Each state is simulated in an isothermal-isobaric ensemble (300K, 1 atm), achieved by the Nosé–Hoover algorithm [35,36] and SSBP. Water molecules are treated as a rigid body. The time step of the MD, Δt, is 0.5 fs.

In each case, 100 ps MD is performed in order to equilibrate the system to 300K and 1 atm. Free energy calculations in both the native and denatured states were carried out in 20 successive stages by changing the coupling parameter λ_i as $i = 0.025, 0.075, \ldots, 0.975$ and by using the double wide sampling [37]. At each λ_i, 5.0 ps MD simulation was performed to equilibrate the system. From the final step of the equilibration, 5.0 ps MD was continued to calculate ΔG. The total 200 ps calculation was performed in both the forward (Ile→Val) and the reverse (Val→Ile) directions. In order to sample conformational space in the vicinity of the initial conformation in NTV, we added mild positional constraints for backbone atoms. The positional constraints were harmonic forces, which become greater as the difference between the instantaneous and the reference coordinates become larger. The force constant was chosen so that the fluctuation of NTV is slightly larger than that of the native state. We have shown that the effect of constraints on the calculated free energy difference is negligible [15].

All the MD/FEP/TI calculations were performed by a newly developed program package based on the framework of Minimization/MD program PRESTO [38].

Acknowledgements

Subroutines for calculating SSBP were kindly provided by Benoît Roux. The X-ray coordinates and the thermodynamic data of the wild-type and mutant human lysozyme were kindly provided by Katsuhide Yutani. Computations were done at the Computer Centers of Kyoto University, Center for Promotion of Computational Science and Engineering of JAERI, Computer Centers of the Institute for Molecular Science, and on the CRAY J916 in our laboratory. This work was supported by grants to A.K. and N.G. from the Ministry of Education, Science and Culture, Japan.

References

1. Kauzmann, W. (1959). Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* **14**, 1-63.
2. Mathews, B.W. (1993). Structural and genetic analysis of protein stability. *Annu. Rev. Biochem.* **62**, 139-160.
3. Serrano, L., Kellis, J.T., Jr, Cann, P., Matouschek, A. & Fersht. A.R. (1992). The folding of an enzyme II. Substructure of barnase and the contribution of different interactions to protein stability. *J. Mol. Biol.* **224**, 783-804.
4. Eriksson, A.E., *et al.*, & Matthew, B.W. (1992). Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science* **255**, 178-183.
5. Buckle, A.M., Henrick, K., Fersht, A.R. (1993). Crystal structural analysis of mutations in the hydrophobic cores of barnase. *J. Mol. Biol.* **234**, 847-860.
6. Takano, K., *et al.*, & Yutani, K. (1995). Contribution of hydrophobic residues to the stability of human lysozyme: calorimetric studies and X-ray structural analysis of the five isoleucine to valine mutants. *J. Mol. Biol.* **254**, 62-76.
7. Dill, K.A. & Shortle, D. (1991). Denatured states of proteins. *Annu. Rev. Biochem.* **60**, 795-825.
8. Evans, P.A., Topping, K.D., Woolfson, D.N. & Dobson, C.M. (1991). Hydrophobic clustering in nonnative states of a protein: interpretation of chemical shifts in NMR spectra of denatured states of lysozyme. *Proteins* **9**, 248-266.
9. Radford, S.E., Buck, M., Topping, K.D., Dobson, C.M. & Evans, P.A. (1992). Hydrogen exchange in native and denatured states of hen egg-white lysozyme. *Proteins* **14**, 237-248.
10. Flanagan, J.M., Kataoka, M., Fujisawa, T. & Engelman, D.M. (1993). Mutations can cause large changes in the conformation of a denatured protein. *Biochemistry* **32**, 10359-10370.
11. Buck, M., Radford, S.E. & Dobson, C.M. (1994). Amide hydrogen exchange in a highly denatured state: hen egg-white lysozyme in urea. *J. Mol. Biol.* **237**, 247-254.
12. Lattman, E.E. (1994). Small angle scattering studies of protein folding. *Curr. Opin. Struct. Biol.* **4**, 87-92.
13. Shortle, D. (1996). The denatured state (the other half of the folding equation) and its role in protein stability. *FASEB J.* **10**, 27-34.
14. Kataoka, M. & Goto, Y. (1996). X-ray solution scattering studies of protein folding. *Fold. Des.* **1**, 107-114.
15. Sugita, Y. & Kitao, A. (1998) Dependence of protein stability on the structure of the denatured state: free energy calculation of I56V mutation in human lysozyme. *Biophys. J.*, in press.
16. Beveridge, B.L. & DiCapua, F.M. (1989). Free energy via molecular simulation: application to chemical and biomolecular systems. *Annu. Rev. Biophys. Chem.* **18**, 431-492.
17. Straatsma, T.P. & McCammon, J.A. (1992). Computational alchemy. *Annu. Rev. Phys. Chem.* **43**, 407-435.
18. van Gunsteren, W.F. & Mark, A.E. (1992). On the interpretation of biochemical data by molecular dynamics computer simulation. *Eur. J. Biochem.* **204**, 947-961.
19. Kollman, P. (1993). Free energy calculations: applications to chemical and biochemical phenomena. *Chem. Rev.* **93**, 2395-2417.
20. Saito, M. (1995). Molecular dynamics/free energy study of a protein in solution with all degrees of freedom and long-range coulomb interactions. *J. Phys. Chem.* **99**, 17043-17048.
21. Tanimura, R. & Saito, M. (1996). Molecular dynamics/free energy perturbation studies of the thermostable V74I mutant of ribonuclease HI. *Mol. Simulation* **16**, 75-85.
22. Tidor, B. & Karplus, M. (1991). Simulation analysis of the stability mutant R96H of T4 lysozyme. *Biochemistry* **30**, 3217-3228.
23. Yun-Yu, S., Mark, A.E., Cun-Xin, W., Fuhua, H., Berendsen, H.J.C. & van Gunsteren, W.F. (1993). Can the stability of protein mutants be predicted by free energy calculations? *Protein Eng.* **6**, 289-295.
24. Saito, M. & Tanimura, R. (1995). Relative melting temperatures of RNase HI mutant proteins from MD simulation/free energy calculations. *Chem. Phys. Lett.* **236**, 156-161.
25. Sun, Y.-C., Veenstra, D.L. & Kollman, P.A. (1996). Free energy calculations of the mutation of Ile96 to Ala in barnase: contribution of the difference in stability. *Protein Eng.* **9**, 273-281.
26. Sugita, Y. & Kitao, A. (1998) Improved protein free energy calculation by more accurate treatment of nonbonded energy: application to chymotrypsin inhibitor 2. *Proteins* **30**, in press.
27. Fiebig, K.M., Schwalbe, H., Buck, M., Smith, L.J. & Dobson, C.M. (1996). Toward a description of the conformations of denatured states of proteins. Comparison of a random coil model with NMR measurements. *J. Phys. Chem.* **100**, 2661-2666.
28. Smith, L.J., Fiebig, K.M., Schwalbe, H. & Dobson, C.M. (1996). The concept of a random coil. Residual structure in peptides and denatured proteins. *Fold. Des.* **1**, 95-106.
29. Smith, L.J., Bolin, K.A., Schwalbe, H., MacArthur, M.W., Thornton, J.M. & Dobson, C.M. (1996). Analysis of mainchain torsion angles in proteins: prediction of NMR coupling constants for native and random coil conformations. *J. Mol. Biol.* **255**, 494-506.
30. Schwalbe, H., *et al.*, & Dobson, C.M. (1997). Structural and dynamical properties of a denatured protein. Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8 M urea. *Biochemistry* **36**, 8977-8991.
31. Weiner, S.J., Kollman, P.A., Nguyen, D.T. & Case, D.A. (1986). An all atom force field for simulations of proteins and nucleic acids. *J. Comp. Chem.* **7**, 230-252.
32. Jorgensen, W.L., Chandrasekhar, J. & Madura, J.D. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926-935.
33. Beglov, D. & Roux, B. (1994). Finite representation of an infinite bulk system: solvent boundary potential for computer simulations. *J. Chem. Phys.* **100**, 9050-9063.

34. Ding, H.-Q., Karasawa, N. & Goddard, W.A., III (1992). Atomic level simulations on million particles: the cell multipole method for Coulomb and London nonbond interactions. *J. Chem. Phys.* **97**, 4309-4315.
35. Hoover, W.G. (1985). Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A* **31**, 1695-1697.
36. Nosé, S. (1984). A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **52**, 255-268.
37. Jorgensen, W.L. & Ravimohan, C. (1985). Monte Carlo simulation of difference in free energy of hydration. *J. Chem. Phys.* **83**, 3050-3054.
38. Morikami, K., Nakai, T., Kidera, A., & Saito, M. (1992). PRESTO (PRotein Engineering SimulaTOr). A vectorized molecular mechanics program for biopolymers. *Comp. Chem.* **16**, 243-248
39. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946-950.
40. Merrit, E.A. & Murphy, M.E.P. (1994). Raster3D version 20. A program for photorealistic molecular graphics. *Acta Crystallogr. D* **50**, 869-873.
41. Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577-2637.

Because *Folding & Design* operates a 'Continuous Publication System' for Research Papers, this paper has been published on the internet before being printed. The paper can be accessed from http://biomednet.com/cbiology/fad – for further information, see the explanation on the contents pages.